



PATENT ABSTRACTS OF JAPAN

(11) Publication number: **11259482 A**(43) Date of publication of application: **24 . 09 . 99**

(51) Int. Cl

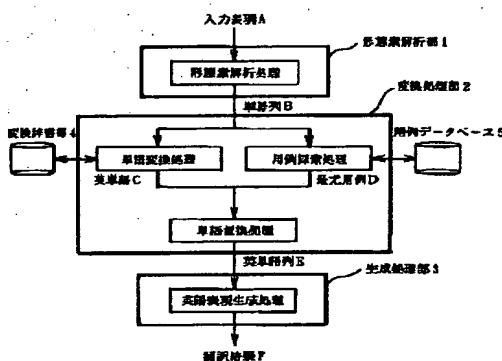
G06F 17/28
G06F 17/22
(21) Application number: **10078542**(22) Date of filing: **12 . 03 . 98**(71) Applicant: **KDD CORP**(72) Inventor: **INOUE NAOKI
SUZUKI MASAMI
HASHIMOTO KAZUO****(54) MACHINE TRANSLATION SYSTEM FOR COMPOSITE NOUN****(57) Abstract:**

PROBLEM TO BE SOLVED: To attain translation without obtaining a syntax structure for an input expression by calculating similarity between an original language expression word and an original language expression example as a sum of similarity between words constituting each expression.

SOLUTION: A morphemic analyzing part 1 divides an input expression A into word units, and outputs a word string B to a conversion processing part 2. The conversion processing part 2 obtains an English word C for each word of the word string B for the input expression A and the original language expression of each sample stored in an example data base, and obtains an example whose similarity is the highest as a maximum likelihood example D. Moreover, the conversion processing part 2 replaces words in the target language expression of the maximum likelihood example D with the preliminarily obtained English words C, and outputs it as an English word string E to a generation processing part 3. The generation processing part 3 generates a correct expression according to English syntax, for example, an (ing) form or a past particle expression

from the inputted English word string E, and outputs a translated result F.

COPYRIGHT: (C)1999,JPO



(19)日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平11-259482

(43)公開日 平成11年(1999)9月24日

(51)Int.Cl.⁶
G 0 6 F 17/28
17/22

識別記号

F I
G 0 6 F 15/38
15/20

P
5 1 2 L
5 2 0 L
5 2 0 M

審査請求 未請求 請求項の数 5 FD (全 9 頁)

(21)出願番号 特願平10-78542

(22)出願日 平成10年(1998)3月12日

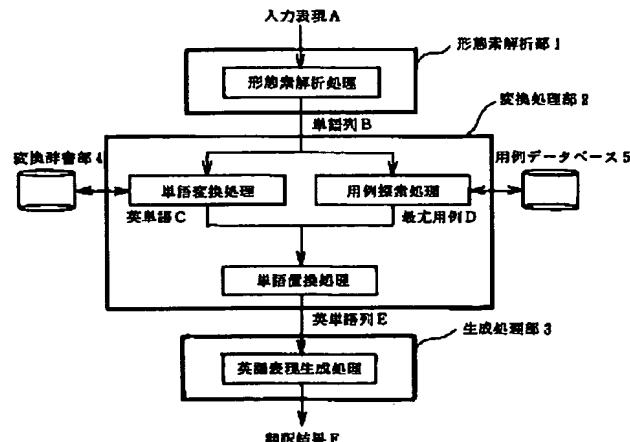
(71)出願人 000001214
ケイディディ株式会社
東京都新宿区西新宿2丁目3番2号
(72)発明者 井ノ上 直己
東京都新宿区西新宿2丁目3番2号 国際
電信電話株式会社内
(72)発明者 鈴木 雅実
東京都新宿区西新宿2丁目3番2号 国際
電信電話株式会社内
(72)発明者 橋本 和夫
東京都新宿区西新宿2丁目3番2号 国際
電信電話株式会社内
(74)代理人 弁理士 大塚 學

(54)【発明の名称】複合名詞の機械翻訳方式

(57)【要約】

【課題】入力表現に対して構文構造を求めることなく翻訳を実現することができる複合名詞の機械翻訳方式を提供する。

【解決手段】原言語(例えば、日本語)表現による複合名詞である入力表現と言語表現による用例が最も類似した最尤用例を求め、入力表現の構文構造に関係なく単語列だけを情報源として、この最尤用例の目的言語(例えば、英語)表現を構成する単語を入力表現に基づいて置き換えることで翻訳を行うことを特徴とする構成を有している。



【特許請求の範囲】

【請求項 1】 原言語表現による複合名詞である入力表現Aは単語単位に分割されて原言語表現単語列Bに変換され、

前記原言語表現単語列B内の各単語に対する目的言語表現への翻訳である目的言語表現単語Cは変換辞書部から求められ、

用例データベースに蓄積された各用例の原言語表現のうちから前記原言語表現単語列Bと最も類似度の高い最尤用例Dが選択され、

該最尤用例Dの原言語表現の各単語を前記目的言語表現単語Cのうちの対応する単語に置換した目的言語表現単語列Eに変換され、

該目的言語表現単語列Eに必要な構文上の修正をして目的言語表現の翻訳結果Fを生成する複合名詞の機械翻訳方式。

【請求項 2】 前記原言語表現単語列と各用例の前記原言語表現との類似度は、それぞれの表現を構成する単語間の類似度の総和として計算されることを特徴とする請求項1に記載の複合名詞の機械翻訳方式。

【請求項 3】 前記原言語表現単語列の単語数と前記の用例の原言語表現の単語数が異なっても、D Pマッチングアルゴリズムを用いて類似度の計算が行われることを特徴とする請求項1に記載の複合名詞の機械翻訳方式。

【請求項 4】 前記原言語表現単語列の単語数と最も類似する用例の前記原言語表現の単語数が異なり、かつ前記の用例の原言語表現の単語数が多い場合には、対応しない単語を前記目的言語表現単語列Eから削除することで前記翻訳結果Fを求めることが特徴とする請求項1に記載の複合名詞の機械翻訳方式。

【請求項 5】 前記用例データベースに蓄積される用例は原言語表現およびその翻訳結果である目的言語表現単語列Eを構成する単語間で予め対応付けが行われていることを特徴とする請求項1に記載の複合名詞の機械翻訳方式。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、機械翻訳方式に関するもので、特に、複合名詞の機械翻訳方式に関するものである。

【0002】

【従来の技術】機械翻訳とは、入力表現を目的言語表現に翻訳することであり、一般には、図6に示すように、入力表現Aに対して、形態素解析処理、構文構造生成処理、構文構造選択処理、単語変換処理、用例探索処理、単語置換処理、英単語列生成処理、英語表現生成処理の8つの処理を、それぞれ形態素解析部11、構文解析部12、変換処理部13、生成処理部14で順々に行い、変換辞書部15、用例データベース16内の情報を参照して目的言語表現へ翻訳している。人間はそれま

で一度も聞いたことがない文を無限に作り出したり理解する能力を持っている。このことは、人間は自分が持っている文法法則を組み合わせることにより今までにない新しい文を作り出したり、また、新しい文を聞いた時も既に持っている文法法則を当てはめて文構造を組み立てこれに基づいて理解していると考えられてきた。構文解析部12において、このような人間と同様の処理を行う目的で入力表現Aの構文構造を求める処理が行われてきた。変換処理部13においては、オンラインドキュメントの増加、計算機の記憶媒体の大容量化、CPUの高速化に伴って、翻訳用例を大量に蓄積した用例データベース16から入力表現Aに最も類似した原言語表現を持つ用例を求め、求めた用例の目的言語表現を構成する単語を入力表現Aに基づいて置き換えることで翻訳を行う方式の研究が進んでおり、翻訳性能は改善されてきている。従来方式では、変換処理部13への入力は入力表現Aの構文構造であったため、出力もそれに対応した英語構文構造Iであった。

【0003】次に、図7に従って従来の翻訳処理の動作を説明する。以下では、依存関係という言葉を用いるが、これは1つの構文構造内の単語間の関係を指しており、依存関係が複数組合わざって1つの構文構造となる。まず形態蓄積解析部11で形態素解析を行い(ステップS11)、単語列Bを生成し、次に構文解析部12で単語間の依存関係を調べて考えられる構文構造候補Gを生成し(ステップS12)、考えられる構文構造候補Gから正しいと思われる最適構文構造Hを選択する(ステップS13)。別の手法として、構文構造の選択は次の変換処理部13における変換処理を行った後に行う手法もある。その際、変換処理は考えられる全ての構文構造に対して行われる。変換処理部13では、求めた最適構文構造Hを構成する単語間の依存関係毎に、依存関係を構成する英単語Cを変換辞書15を参照して求め(ステップS14)、依存関係の表現と最も類似した最尤用例Dをデータベース16を参照して求める(ステップS15)。さらに、求めた最尤用例Dの英語表現をステップS14で求めた英単語Cで適切に置き換えて、依存関係に対する英語表現を作成する(ステップS16)。上記ステップS14からステップS16までをすべての依存関係に対して繰り返し、生成処理部14で入力表現Aに対する英語構文構造Iから英単語列Eを生成した(ステップS17)後、複数形やing構文など英語表現として適切な表現を作成して翻訳結果Fとして出力する(ステップS18)。

【0004】

【発明が解決しようとする課題】即ち、従来は、何らかの方式により入力表現Aの構文構造を求める必要があった。そのため、構文構造を正しく求めることができなかった場合、翻訳誤りが生じることになる。構文構造は、一般に複数個が存在し、しかも全ての構文構造の中から

正しい構文構造を選択することは困難であるため、決定された構文構造が誤っていることは頻繁に生じていた。このように、従来の機械翻訳方式では、解析処理において入力表現の構文構造を1つに決定していたため、翻訳性能の劣化が生じている。これは、複合名詞の機械翻訳方式においても同様である。

【0005】本発明は、入力表現に対して構文構造を求めることなく翻訳を実現することができる複合名詞の機械翻訳方式を提供することを目的とする。

【0006】

【課題を解決するための手段】この目的を達成するためには、本発明による複合名詞の機械翻訳方式は、原言語表現による複合名詞である入力表現Aは単語単位に分割されて原言語表現単語列Bに変換され、前記原言語表現単語列B内の各単語に対する目的言語表現への翻訳である目的言語表現単語Cは変換辞書部から求められ、用例データベースに蓄積された各用例の原言語表現のうちから前記原言語表現単語列Bと最も類似度の高い最尤用例Dが選択され、該最尤用例Dの原言語表現の各単語を前記目的言語表現単語Cのうちの対応する単語に置換した目的言語表現単語列Eに変換され、該目的言語表現単語列Eに必要な構文上の修正をして目的言語表現の翻訳結果Fを生成することを特徴とする構成を有している。

【0007】

【発明の実施の形態】上記目的を達成する本発明の第1の形態に係る機械翻訳方式は、原言語表現単語列Bと原言語表現用例間の類似度をそれぞれの表現を構成する単語間の類似度の総和として計算することを特徴とする。上記目的を達成する本発明の第2の形態に係る機械翻訳方式は、D Pマッチングアルゴリズムを用いることで原言語表現単語列Bの単語数と原言語表現用例の単語数が異なっても、類似度の計算が行えることを特徴とする。上記目的を達成する本発明の第3の形態に係る機械翻訳方式は、原言語表現単語列Bと最も類似する目的言語表現単語列Eを構成する単語から、前記原言語表現単語列Bにない単語を削除することを特徴とする。上記目的を達成する本発明の第4の形態に係る機械翻訳方式は、原言語表現およびその翻訳結果である目的言語表現単語列Eを構成する単語間の対応付けが予め行われて用例データベースに蓄積されていることを特徴とする。

【0008】

【実施例】本発明による複合名詞の機械翻訳方式は、図1に示すように、形態素解析部1、変換処理部2および生成処理部3により構成される。形態素解析部1は、入力表現Aを単語単位に分割し、単語列Bを変換処理部2へ出力する。ここで、単語列Bを構成する各単語には、品詞などの単語に関する属性情報が含まれている。変換処理部2は、まず、単語列Bの各単語に対する英単語Cは変換辞書部4を探索して求める。次に、入力表現Aに対する単語列Bと、用例データベース5に蓄積された各

用例の原言語表現との類似度を計算し、最も類似度の高い用例を最尤用例Dとして求める。さらに、変換処理部2は、最尤用例Dの目的言語表現内の単語を先に求めておいた英単語Cと置換し、英単語列Eとして生成処理部3へ出力する。上記置換において、用例の原言語表現内に入力表現A中の単語と対応する単語が無い場合、その単語に対する用例の目的言語表現内の単語をnullという特別な単語に置換する。生成処理部3は、入力された英単語列Eから、例えば前記nullを削除したり、ing形や過去分詞表現にするなど英語構文上正しい表現を生成して翻訳結果Fを出力する。

【0009】前記変換処理部2において、最尤用例Dを求めるに際して、入力表現Aおよび用例の原言語表現を構成する単語間の類似度の総和としてこれら表現間の類似度を求める。最尤用例Dはこの類似度の最も高い用例である。例えば、入力表現Aに対する単語列Bが「日本橋」「一丁目」「略図」であり、原言語表現の用例の単語列Bが「浅草」「一丁目」「略図」である場合、これら単語列間の類似度は「日本橋」と「浅草」間の類似度、「一丁目」と「一丁目」間の類似度、「略図」と「略図」間の類似度の総和として求める。

【0010】また、前記変換処理部2において、最尤用例Dを求めるに際して、入力表現の単語数と原言語表現の用例の単語数が異なる場合でも、考えられる全ての単語間の対応の中で最も高い類似度を求める。例えば、入力表現Aの単語列Bが「日本橋」「一丁目」「略図」と3単語からなり、用例の原言語表現の単語列Bが「銀座」「一丁目」「通り」「略図」と4単語からなっている場合、考えられ単語間の対応は図2に示すように、(「日本橋」-「銀座」, 「一丁目」-「一丁目」, なし-「通り」, 「略図」-「略図」という対応パターン1の他に、(「日本橋」-「銀座」, なし-「一丁目」, 「一丁目」-「通り」, 「略図」-「略図」の対応パターン2), (「日本橋」-「銀座」, 「一丁目」-「一丁目」, 「略図」-「通り」, なし-「略図」の対応パターン3)及び(なし-「銀座」, 「日本橋」-「一丁目」, 「一丁目」-「通り」, 「略図」-「略図」の対応パターン4)という全部で4通りの対応が考えられる。入力表現と用例の原言語表現間の類似度は、これら全ての対応に対する単語間の類似度の総和の中で最も高い類似度として求める。ここで、単語間の対応が交さすることは考慮外である。

【0011】前記変換処理部2において、入力表現Aとの類似度を計算する用例とは、原言語表現とその翻訳である目的言語表現とが単語単位で対応付けされたデータのことである。例えば、「銀座一丁目通り略図」とその翻訳である“View of GinzaDoori 1-choume”は図3に示すように、日本語の「銀座」が英語の“Ginza”に対応し、さらに「一丁目」が“1-choume”に、「通り」が“Doori”に、「略図」が“View”に対応付けされてい

る。また、英語の“of”に対応する日本語がないことも示されている。

【0012】本発明の機械翻訳方式は、図4に示すフローチャートに従い、次のように実施される。まず、形態素解析部1に日本語表現Aが入力されると形態素解析を行い単語列Bを出力する(ステップS1)。次に変換処理部2では、単語列Bの全ての単語に対して変換辞書部4を参照して英単語Cを求める(ステップS2)。さらに、変換処理部2では、用例データベース5を参照して単語列Bと最も類似した最尤用例Dを求める(ステップS3)、引き続き、求めた最尤用例Dの原言語表現の単語列と入力表現Aの単語列Bとの対応を求める(ステップS4)。ここで、前記の単語列Bと用例の原言語表現との類似度の計算において、DP(Dynamic Programming)マッチングアルゴリズムを用いて類似度を計算する。D

$$S(i, j) = \max \begin{cases} S(i-1, j) + s(x_i, 0) \\ S(i-1, j-1) + s(x_i, y_j) \\ S(i, j-1) + s(0, y_j) \end{cases} \quad (式1)$$

【0015】ここで、 $s(x_i, y_j)$ は単語 x_i と y_j の間の類似度であり、 $s(x_i, 0)$ は単語 x_i が削除されるコスト、 $s(0, y_j)$ は単語 y_j が挿入されるコストである。同様に、 $j = 2, 3, 4$ に対して $i = 0, 1, 2, 3$ の順に $S(i, j)$ を求め、最終的に $S(3, 4)$ が求める2つの単語列間の類似度となる。式1は繰り返し関数の形をしているため、類似度 $S(i, j)$ は格子点 $(0, 0)$ から (i, j) までの全てのパスの中で最も高い類似度となる。 $S(3, 4)$ を計算した後、格子点 $(3, 4)$ に至った1つ手前の格子点(図5では、格子点 $(2, 3)$)を順次 $(0, 0)$ まで遡ることで、格子点 $(0, 0)$ から $(3, 4)$ までのパスがただ一つ求まり、そのパスが類似度 $S(3, 4)$ を得る時の単語間の対応を表す。

【0016】例えば、格子点 $(3, 4)$ に至るパスが図5の通りであった場合、縦軸に配列された単語列「日本橋」「一丁目」「略図」と横軸に配列された単語列「銀座」「一丁目」「通り」「略図」の間の全ての対応の中で最も類似度の高い対応は「日本橋」-「銀座」、「一丁目」-「一丁目」、「略図」-「略図」であることを表している。このように、DPマッチングアルゴリズムを用いることにより、最も類似度の高い単語間の対応関係とその時の類似度を求めることができる。

【0017】さらに、変換処理部2は、前記ステップS3で求めた最尤用例Dの目的言語表現をステップS2で求めた英単語で適切に置き換える(ステップS5)ことを単語列Bの全ての単語に対して繰り返す。次に最尤用例Dの目的言語表現中に入力表現Aに対応する単語がない単語はnullという特別な単語に置換し(ステップS6)、英語単語列Eを作成する。このnullという単語は、後の生成処理部3において削除されることになる。例えば、入力表現Aが「日本橋一丁目略図」で、用例デ

Pマッチングアルゴリズムを用いることにより、単語列Bと用例の原言語表現との対応関係も求めることができる。

【0013】例えば、単語列Bが「日本橋」「一丁目」「略図」の場合の実施例を図5に従い説明する。図5では、単語「日本橋」、「一丁目」、「略図」は x_i ($i = 1, 2, 3$)で示され、別の単語列「銀座」「一丁目」「通り」「略図」の単語は y_j ($j = 1, 2, 3, 4$)で示されている。まず、図5中の格子点 $(i, 0)$ ($i = 0, 1, 2, 3$)における類似度 $S(i, 0)$ を初期値0に設定する。次に、 $j = 1$ に対し、 $i = 0, 1, 2, 3$ の順に $S(i, j)$ を式1に従って求める。

【0014】

【数1】

$$S(i, j) = \max \begin{cases} S(i-1, j) + s(x_i, 0) \\ S(i-1, j-1) + s(x_i, y_j) \\ S(i, j-1) + s(0, y_j) \end{cases} \quad (式1)$$

ータベース5中の用例のうち最も言語表現が類似した用例が図3に示した「銀座一丁目通り略図」-“View of Ginza Doori 1-choume”であり、単語列Bと用例の原言語表現との対応が「日本橋」-「銀座」、「一丁目」-「一丁目」、「略図」-「略図」であった場合、単語「銀座」に対する英語“Ginza”を単語列Bの単語「日本橋」の英語“Nihonbashi”で置き換える。同様の処理を用例の原言語表現の単語「一丁目」「通り」「略図」に対して行うが、ここで、「通り」に対する入力表現中の単語がないので、英語“Doori”に変わってnullという特別な単語を代入する。その結果、入力表現「日本橋一丁目略図」に対する変換結果として“View of Nihonbashi null 1-choume”を英単語列Eとして得る。

【0018】次に、生成処理部3では、変換処理部2からの出力である英単語列Eに対し、英語表現として正しい表現を翻訳結果Fとして出力する。上記の例では英単語列Eからnullを削除して、“View of Nihonbashi 1-choume”を翻訳結果Fとして出力する。

【0019】

【発明の効果】以上、実施例に基づいて具体的に説明したように、本発明では、入力表現と用例の原言語表現との類似度を計算する段階で、DPマッチングアルゴリズムを用いて入力表現の単語列のみから計算しているので、入力表現の構文構造を求める必要がない。本発明を、原言語表現と目的言語表現の対応関係を前述した用例を蓄積した用例データベースから、入力表現と用例の原言語表現とが最も類似した用例を求める、この用例の目的言語表現を構成する単語を入力表現に対する英単語で置き換えることで翻訳を行う機械翻訳方式に用いることにより、翻訳性能の向上を期待することができる。

【図面の簡単な説明】

【図1】本発明を用いた機械翻訳方式の構成例図であ

る。

【図2】入力表現と用例の原言語表現との対応関係の一例を示す図である。

【図3】本発明において用例データベースに蓄積された用例を示す図である。

【図4】本発明の実施例を説明するためのフローチャートである。

【図5】本発明に用いるDPマッチングアルゴリズムを説明するための図である。

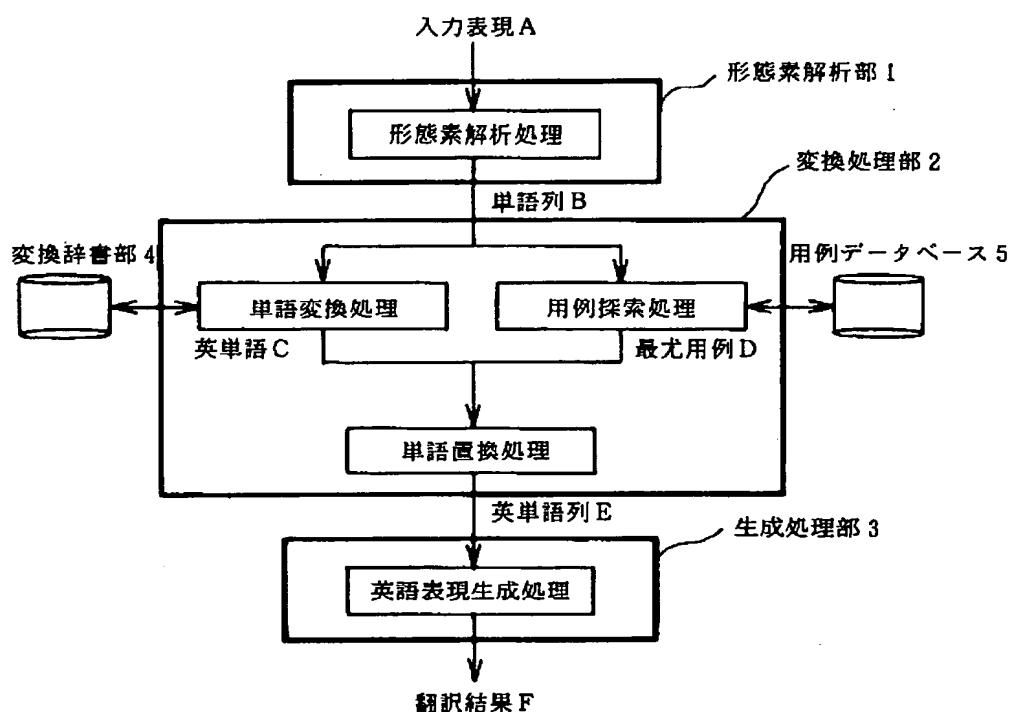
【図6】従来の機械翻訳方式例を示すブロック図である。

【図7】図6の従来例の動作を説明するためのフローチャートである。

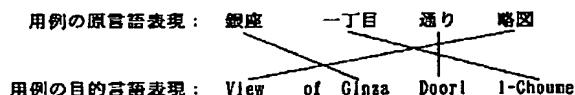
【符号の説明】

- 1 形態素解析部
- 2 変換処理部
- 3 生成処理部
- 4 変換辞書部
- 5 用例データベース
- 1 1 形態素解析部
- 1 2 構文解析部
- 1 3 変換処理部
- 1 4 生成処理部
- 1 5 変換辞書部
- 1 6 用例データベース

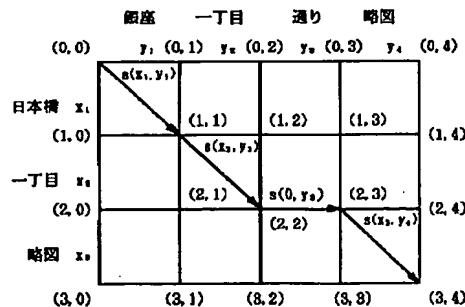
【図1】



【図3】

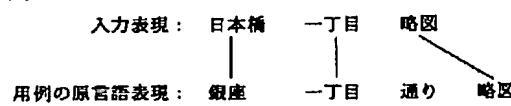


【図5】

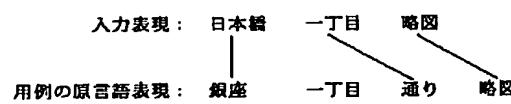


【図2】

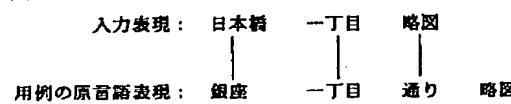
〔対応パターン1〕



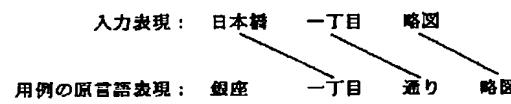
〔対応パターン2〕



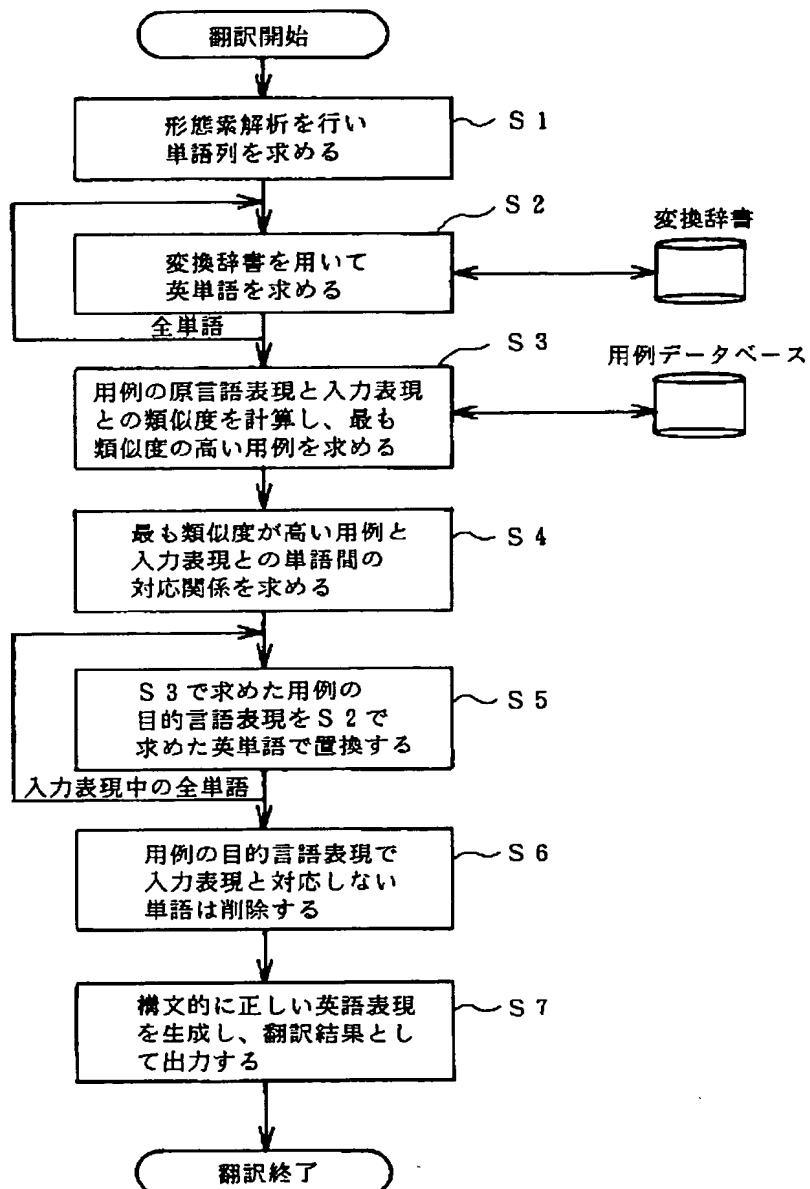
〔対応パターン3〕



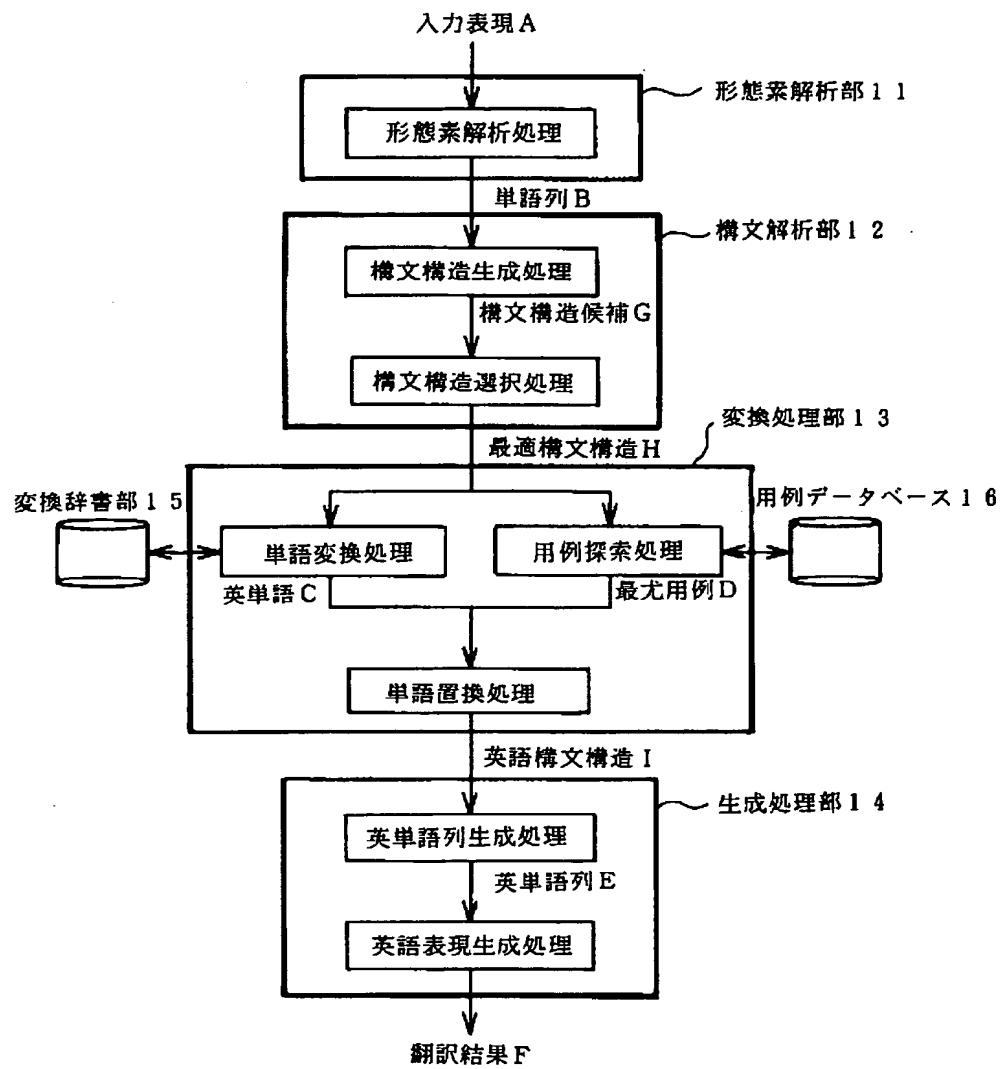
〔対応パターン4〕



【図4】



【図6】



【図7】

